DOCUMENT RESUME

ED 243 425 IR 011 034

AUTHOR King, F. J.; Roblyer, M. D.
TITLE Practical Strategies for Encouraging Research in Computer-Based Instruction.
PUB DATE Jan 84
NOTE 30p.; Paper presented at the Annual Meeting of the Association for Educational Communications and Technology (Dallas, TX, January 20-24, 1984). For proceedings, see IR 011 020.
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computer Assisted Instruction; *Computer Software; Flow Charts; *Models; Regression (Statistics); *Research Design; *Research Methodology; *Statistical Analysis

ABSTRACT

This paper presents five designs (three one-group and two two-group) which can be effectively employed to study computer-based methods in non-laboratory settings, including actual classroom implementation, where a non-treatment control design may not be feasible. These designs are sequential analysis, value-added analysis, non-equivalent dependent variables, regression-discontinuity, and cohort design. For each design, examples of research studies which effectively employ such designs are given. Sample utility programs and procedures are provided which can be used by persons with limited training in statistics and research methods to analyze data resulting from these non-traditional designs. A decision flowchart is included, as well as a description of how to select the most appropriate non-traditional design for a specific project. Three appendices provide an Applesoft BASIC program for sequential analysis; computations for value-added analysis; and analysis and results of the regression-discontinuity design. Twenty-two references are listed. (LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243425

PRACTICAL STRATEGIES FOR ENCOURAGING RESEARCH IN COMPUTER-BASED INSTRUCTION

Paper presentation for the Annual Conference of the Association for Educational Communications and Technology
Dallas, Texas
January 21-24, 1984

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

F. J. King, Florida State University
M. D. Roblyer, Florida A & M University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Michael Simonson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

INTRODUCTION

The widespread use of microcomputers in schools which began in the first part of this decade signalled a new era for educational computing research. Before this movement, most studies of computer-based instruction were limited to large-scale, centralized projects such as the Stanford math and reading (Fletcher & Atkinson, 1972; Suppes & Morningstar, 1972), Computer Curriculum Corporation, (Lysiak, Wallace & Evans, 1976; Crandall, 1977; Holland, 1980), Control Data PLATO (Alessi, Siegel, Silver & Baines, 1982-83; Poore, Qualls & Brown, 1981), and TICCIT (Jones, 1978) programs. These studies were usually performed by grant-funded projects with a research and evaluation component built into them. But with the acceptance of low-cost computing in schools and the explosion of new computer-based materials, the opportunities for research in this area have greatly expanded and changed in nature. Small school and classroom-based programs have been started throughout the country. Not only do studies of these programs have the potential for making a significant contribution to the field, they are, in fact, essential if we are to make any progress in establishing the usefulness of new computer-based methods and materials.

However, since the focus of these educational computing projects is often implementation rather than research, at least two factors are likely to inhibit systematic studies of their activities. First, there are practical difficulties in arranging research studies with the traditional randomized two-group designs in these settings. Secondly, project personnel usually lack the expertise to employ other effective designs and analyze the resulting data. If we are to take advantage of the valuable information available from classroom computer projects, alternative designs to study computer-based activities must be employed, and "cookbook" data analysis methods for each design must be made widely available. This paper will:

- Present five designs (three one-group and two two-group) which can be effectively employed to study computer-based methods when a randomized two-group experimental design is not practical,

- For each design, give examples of research studies which effectively employ such designs,

- Supply example utility programs and procedures which can

IR011034

be used by persons with limited training in statistics and research methods to analyze data resulting from these non-traditional designs, and finally,

- Give a decision flowchart and describe how to select the most appropriate non-traditional design for one's project.

FIVE NON-TRADITIONAL RESEARCH DESIGNS

Design #1: Sequential Analysis

Sequential analysis techniques originated almost forty years ago (Wald, 1947), but, with few exceptions, they have been used for quality control of products rather than to analyze results of research studies. However, they can also be used to construct an effective behavioral research design. In such studies, all sequential analysis techniques would employ the same strategy: (a) an observation would be made (e.g. one or more students are tested on some skills), (b) the results are recorded on a chart or graph developed for this purpose, and (c) by reading the chart, a decision is made to accept the null hypothesis, to accept an alternative hypothesis or to make another observation.

This technique has many advantages for computer-based studies besides not requiring a control group. Its principal advantage is efficiency. As a sampling technique, it can allow decisions to be made with as much as 50% fewer observations than would be required for procedures in which sample size must be specified in advance. This makes it an ideal choice when the number of people which can be tested on the materials is limited by factors such as the number of computers available or by constraints such as not being able to get the whole student group at one time or the expense of the treatment.

There are some limitations on when it can be effectively used. Since there is no control in this design for prior learning, it is better to use it when there is little likelihood that students would be able to do the skills involved before exposure to the materials, as with programming concepts or verbal information drill such as in social studies or history. It is also desirable that the skill units be relatively short, to cut down on effects due to maturation and history (Campbell and Stanley, 1963). If the materials under study are lengthy, they can be divided up into short units, and a sequential analysis can be done on each one. This technique is also most useful when teachers want to make an absolute judgement about materials ("Are these computer-based lessons effective?") rather than a comparative one ("Are these lessons better than non-computer lessons?")

Example of sequential analysis use. A typical study using the sequential analysis method is currently being undertaken by the U. S. Army to evaluate the effectiveness of microcomputer/videodisc training materials. The modules were designed to teach specialist students to perform a lengthy and

expensive procedure for testing the purity of jet fuel. No other training is currently being given, so no control condition is available. Effectiveness is to be defined in terms of the proportion of students who successfully complete the simulated task in no more than two trials. Based on the experience of the faculty of the school in training similar students on similar tasks, that proportion is determined to be at least .80. After this proportion is set, the following steps are to be completed:

STEP 1: Setting acceptable/unacceptable quality levels - Since absolute accuracy in judging instructional quality cannot be achieved with a sampling technique, the success proportion (.80) must be translated into a range of proportions which define acceptable and unacceptable instructional quality. In this instance, the range is set at .80 plus or minus .10. The Acceptable Quality Level (AQL) or the proportion of the population which indicates acceptable quality, is .90. If 90% or more of the students in the population can meet te success criterion, the instruction is considered to be of high quality and accepted for routine use. The Rejectable Quality Level (RQL) or the proportion of the population that indicates unacceptably low quality, is .70. This means that if only 70% or fewer of the students can meet the success criterion, the instructional is rejected as unacceptable. If the proportion alls between .70 and .90, the instruction is of questionable quality.

STEP 2: Setting risk parameters and proportions - The probablity of rejecting high quality instruction (alpha) and the probablity of accepting low quality instruction (beta) must be set. Because obtaining high quality instruction is very difficult, the probability of rejecting wrongly it must be very low, thus alpha is set at .01. Mistakenly accepting low quality instruction for quality instruction is not deemed to be very serious because it can always be remedied later, thus Beta is set at .10. One more calculation is necessary. Since sequential analysis was originally designed to detect proportions of defective units, p(1) and p(2) must be calculated, respectively, as 1-AQL and 1-RQL.

STEP 3: Construct Operating Characteristics Curve (OC), Average Sample Size Curve (ASN) and truncation number - Now we have all the values to construct (a) an OC Curve, which will show the probability of rejecting or accepting the null hypothesis for any true value of the success proportion under consideration (given the specifications of the test), (b) the ASN, which gives the average number of observations needed to reach a decision for any true value of the success proportion and (c) the truncation number, which is the largest sample size that could be needed to make a decision without altering the risks the experimenter is willing to accept. Figures 1 and 2 illustrate the OC and ASN curves with sample data. The truncation number is 113. The slope and intercepts for the lines in the figures, as well as the truncation number, are calculated from the values derived or set in Step 1 above. (For further directions on how to calculate values, see Wald (1947), Burr (1980) or Epstein (1979).

[Figures 1 and 2 about here.]

Figure 1 (the OC) shows that, for very high quality instruction ($p < .10$), the probability is almost 1.0 that it will be accepted (i.e., the null hypothesis that $p > .30$ will be rejected). For very low quality instruction ($p < .30$) the probability of rejecting the null hypothesis (i.e., accepting the instruction) drops from .10 to almost zero as p increases. For values of p between .30 and .10, the probability of rejecting the null hypothesis varies from .99 to .10. If the population proportion is .20 (1-.80), for example, the probability of rejecting the null hypothesis is about .64.

Figure 2 (the ASN curve) shows that the largest average number of observations (38) required for a decision to be made occurs when the population proportion (p) is equal to .20. As p varies on either side of .20, the ASN decreases. The truncation number represents the worst possible case in which 113 subjects would have to be observed before the null hypothesis would be accepted or rejected.

STEP 4: Construct graph for recording observations. The graph for recording observations contains three zones corresponding to the three possible decisions which are considered each time an observation is made (to reject, accept, or take more samples). The actual decision choice is determined by the pattern of recorded data. A graph of sample data is shown in Figure 3.

[Figure 3 about here.]

The data were placed on the graph in the following way. The first student is tested and if the student passes, a line is drawn horizontally from 0 to 1. If the student fails, a horizontal line is constructed from 0 to 1 and a vertical line drawn from 1 to 1. Both points are in the zone labeled "continue sampling," so another student is tested and the results placed on the graph in the same way. This process continues until either the bottom line is crossed and the decision is made to accept the instruction (reject the null hypothesis), or the upper diagonal line is crossed and the decision is made to reject the instruction (accept the null hypothesis). If sampling continued until the truncation number was reached, the instruction would be accepted if the last point were nearer the bottom diagonal than the top diagonal, or it would be rejected if the opposite prevailed. In Figure 3, the first eight students passed, the ninth failed, the next six passed, and the decision was made to

accept the instruction.

Program to do sequential analysis calculations. A utility program to do the necessary calculations mentioned in Step 3 above is given in Appendix A. The experimenter must enter AQL, RQL, alpha and Beta. The program can produce:

1) The intercepts and slope for both the upper and lower lines of the acceptance/rejection regions.
2) The truncation number
3) A table of values of the coordinates to plot an OC.
4) A table of values of coordinates to plot an ASN curve.

Design #2: Value-added Analysis

The basic idea underlying this model is that the effect of a treatment can be estimated by comparing the average observed growth between pretest and posttest with the estimated growth which is expected to occur in the absence of a treatment (Bryk & Weisberg, 1976). An unbiased estimate of the estimated growth rate can be provided via ordinary least squares regression of pre-test on the subject's age, if age at the time of pretest is not related to systematic growth. When growth (estimated from a cross-section of the total population) is not linear, transformations must be made before value-added analysis can be completed.

Many computer-based treatments deal with such areas as reading and writing skills, motor skills development and increasing problem-solving ability through use of tools like LOGO. Study of such treatments is difficult because treatment must extend over a long period of time in order for measurable differences to occur, giving rise to the possibility of effects due to maturation and history. The value-added analysis is ideal for research into these long-term activities since it controls for these effects.

Certain conditions must, however, be present. An independent variable which can be measured without error (i.e. chronological age) must be available, and it must be correlated with pretest scores. Also, a large enough number of students must be available to be able to do a product-moment correlation.

Example of value-added analysis use. This procedure was used in a study of a microcomputer-based reading and writing program in grades K-1 (Garretson, Vertuno, King and Roblyer, 1982). All students in these grades (approximately 100) were to receive the computer-based instruction, which was to take place over the school year. Several measures were used to measure skill levels, including a Nonsense Words Test devised by the school resource teacher, and several subtests of the California Test of Basic Skills. The value-added analysis was used with the Nonsense Words Test. (Analysis methods for the other test will be discussed later under another design.) The following steps were completed:

STEP 1: Estimate of growth rate. First, an unbiased

estimate of growth (G) was obtained. This is done by regressing pretest scores on age and using the resulting regression coefficient (b) as the estimate of growth rate, then using the following formula:

$$G = \bar{y}_1 + b(a_2 - a_1)$$

where $\bar{y}(1)$ is the pre test mean and a(2) and a(1) are the ages at pre and post test times.

STEP 2: Calculate value added. Use the following formula to determine the effect of the treatment, taking into account the natural growth of students over time:

$$V = \bar{y}_2 - G$$

where $\bar{y}(2)$ is the post test mean.

Unfortunately, no test of the significance of V exists. However, Bryk et al. (1980) pointed out that the jack knife technique (Mosteller and Tukey, 1977) can be used to provide a test statistic. It involves the computation of a pseudo-value (V*i) for each individual in the sample, treating these values as data points, and calculating their mean and standard errors. The mean is an unbiased estimate of V and the standard error allows the calculation of a t-test (df = N-1) for significance testing or interval estimation. In the K-1 study, this yielded the following results:

$$V = 35.3 - 21.3 = 14.00$$

$$t \text{ ratio} = 5.36 \qquad p < .05$$

To increase the precision of this model, Bryk et al (1980) also describe how it can be extended to incorporate background variables that may be related to individual growth rates. This is done by regressing the pre test on age and the first order interactions of age and each background variable.

SPSS procedure to do calculations. The listing of an SPSS program to do the above calculations is shown in Appendix B.

Design #3: Non-equivalent Dependent Variables

This design involves one group of subjects to which two (or more) variables are administered as pretests. Performance on one of the variables is expected to change as a result of a treatment, while the others are not expected to change (Cook and Campbell, 1979). Such an approach has two advantages. First, no non-treatment control group is needed, since performance on one variable acts as a control for the other. Second, older students can be used since, unlike the value-added analysis, no independent variable such as age is required. The basic design is diagrammed as:

```
        Ø        X        Ø
       1A        A       2A
- - - - - - - - - - - - - - - - - -
        Ø                 Ø
       1B                2B
```

where Ø(1A) and Ø(2A) are pre and post tests for the variable that is hypothesized to be affected by the treatment and Ø(1B) and Ø(2B) are pre and post tests for the variable hypothesized not to be affected.

In order to be sure this design will be credible for the situation, the researcher must (a) specify in advance which variables are expected to change and which are not; (b) demonstrate that differential change is not due to differential reliability of measures or to ceiling/basement effects; and (c) use variables which are conceptually so similar that both would be equally affected by the same threats to internal validity, namely maturation and testing.

For example, suppose a school wanted to evaluate a computer-based module on birds and mammals, and they also had available a test on reptiles. Students could be given pretests on both units, instruction on the birds and mammals units only, and then post tests on both units. If statistically and educationally significant changes occur on the birds and mammals unit but not on the reptiles test, the birds and mammals unit is judged to be effective. Of course, they would have to be satisfied that the reptiles test was not more unreliable that the other one, and that students were not already scoring at the ceiling of the reptiles test at the time they were pretested. In addition, the investigators would have to be satisfied that the change in the bird and mammals test was not due to some learning experience other than the modules, such as a television show about this topic during the treatment period.

Cook and Campbell (1979) point out that this basic design is weak because it is hard to be certain that all conditions necessary for its validity have been met. However, it can be made much stronger by increasing the number of measurement "waves" or the number of variables. This also presents an excellent design for

evaluating two or more modules at the same time. In this expanded design, instruction on one variable is given, followed by testing on all variables. Then instruction on the next module is given, followed by testing on all variables, and so on. This design may be shown as:

```
        Ø        X        Ø
       1A        A       2A
- - - - - - - - - - - - - - - - - - - - - - - - - - - - etc.
        Ø                 Ø        X        Ø
       1B                2B        B       3B
```

If the expected pattern of change occurred, the conclusion concerning the effectiveness of both units would be relatively strong. If the expected change between Ø(1A) and Ø(2A) occurred but the expected change between Ø(2B) and Ø(3B) did not occur, the conclusions about the effectiveness of either unit would be uncertain. Steps in analyzing data from such a design are as follows:

STEP 1: Determine pre-post differences. Means are determined for each groups' test results and a difference calculated between each:

$D_1 = Ø_{2A} - Ø_{1A}$ $D_2 = Ø_{2B} - Ø_{1B}$ $D_3 = Ø_{3B} - Ø_{2B}$

STEP 2: Determine significance of differences. A t-test is done between each pre and post test pair to determine if post test scores were significantly different from pre test scores.

STEP 3: Compare results or groups. If D(1) and D(3) are significantly significant and D2 is not, the instruction for the two modules under consideration can be said to be effective.

Berquist and Graham (1980) expanded this design to allow the evaluation of instruction of many objectives over many waves of measurement. Their design is shown in Figure 4 below. They point out that this approach is useful to evaluate instruction with single students as well as groups.

[Figure 4 about here.]

If more than one unit of instruction is to be evaluated and the group is sufficiently large (e.g. 60 students), this design can be converted to a powerful true experimental design which does not require a non-CAI control group. Where instruction can be individually administered to each student (as with CAI), the

following design could be used:

R	Ø 1A	Ø 1B	X A	Ø 2A	Ø 2B
R	Ø 1A	Ø 1B	X B	Ø 2A	Ø 2B

Here, students must be randomly assigned (R) to treatment groups (Xa or Xb) and pre and post tested with both unit tests (Ø1A and Ø1B).

Example of true experimental design involving non-equivalent dependent measures. One study using this design as part of its total evaluation program was the ETS/LAUSD study of computer-assisted instruction and compensatory education (Ragosta, Holland and Jamison, 1982). Schools in the Los Angeles Area School District used math and language CAI available from the Computer Curriculum Corporation. In Grade 4 students were randomly assigned to receive two sessions of CAI daily. They received either: (a) two sessions of mathematics (MM), (b) one session of reading and one of language arts (RL, or (c) one session of mathematics with one session of reading or language arts on alternate days (MRL). All students were pretested in the fall with the Iowa Test of Basic Skills (ITBS) and a curriculum-specific test (CST) in mathematics, language and reading. They were posttested the next spring with the California Test of Basic Skills (CTBS) and the CST. Thus, the RL students were controls for evaluating the effectiveness of two levels of mathematics instruction (MM and MRL), while the MM students were controls for the evaluation of two levels of reading (RL and MRL). A regression analysis was used to analyze the data. These analysis procedures are the same as those used in Design #4 which follows.

This design is really better in some ways than one having a non-CAI control group because it is a better control for Hawthorne effects. The novelty of the CAI experience is the same for both groups so that the computer aspect alone cannot be expected to produce significant achievement differences between groups.

Design #4: Regression-Discontinuity

Although this is the one design which actually calls for a no-treatment control group, the regression-discontinuity design is especially useful when compensatory or enrichment programs are to be evaluated. This is so because of its requirement that students be pretested and all students below (or above) a certain percentile be placed in a treatment group All those above (or below) the cutting score (i.e. those who would not be eligible for the program anyway) comprise the control group. At the end of the treatment time, the postest is administered to all those who received the pretest, and a "dummy" treatment variable is

constructed by assigning a score of "1" to each member of the treatment group and a "0" to each member of the "non-treatment group." Finally, a regression analysis is done using the following model:

$$\hat{Y} = a + b_1 T + b_2 Pre$$

where $\hat{Y}$ is the adjusted posttest score, T and Pre are the treatment and pretest variables, a is the regression intercept, b(1) is the difference between the adjusted means of the treatment groups, and b(2) is the regression coefficient for the pretest. If b(1) is statistically significant, the treatment is effective.

The principal advantage of the regression-discontinuity design over other quasi-experimental designs is that it is not biased by errors of measurement in the pretest. That is, in the absence of a treatment effect, the expected posttest adjusted mean difference is zero. That is not true in designs in which unknown selection variables operate in the formation of treatment and control groups.

A primary disadvantage of the design when compared to a randomized treatment-control group design is that it has low power in being able to reject the hypothesis of no differences between groups. If a randomized design used a sample of 100 subjects, the regression-discontinuity design would require 275 subjects to have equal power (Reichardt, 1979). In addition, if one of the groups is relatively small, the estimate of the regression of post test on pre test for that group may be unstable.

Two other precautions should be kept in mind when using this design. First, the process of assigning students to groups should be done without error. If the pretest is to be used as the covariate in the analysis, selection should be based on it only. If other sources of information, such as teacher or parent judgements, are used in addition to or instead of the pretest, the analysis may be seriously biased. Huitema (1980) gives procedures for dealing with multiple sources of information in the assignment process. The second precaution in using this design is that non-linear relationships between pre and post tests, which result from selection-motivation factors, may cause spurious treatment effects to occur. Reichardt (1979) suggested that the first step in analyzing the data from this design should be to plot the raw data to see what relationships are suggested. If non-linear or interactive effects are suggested, the analytic model given previously can be expanded in an attempt to account for them.

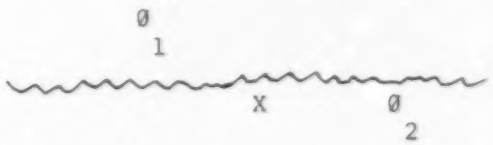
Example of regression-discontinuity use. Although no use of this design is available from the literature, an example of a data file, an SPSS procedure file and the results from this

procedure are given in Appendix D, along with steps for completing the procedure.

Design #5: Cohort Design

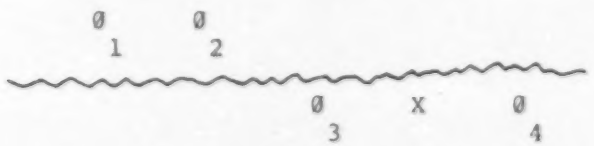
Cohorts are groups of persons who follow each other through an institution. For example, last year's first graders are cohorts of this year's first graders. The cohort design compares the performance of students in a treatment group with their cohorts that did not receive the treatment. This design is especially useful in evaluating the effectiveness of newly-installed, year-long computer-based programs where it is not feasible to withhold the program from some students in order to have a control group. The design makes the assumption that student cohorts are comparable in most respects, and that differences between them on a dependent variable can be attributed to a treatment administered to one of the groups.

In its simplest form, the design is as follows:



Ø(1) represents, for example, a final achievement measure of students in a certain grade in a given year, and Ø(2) represents the same final achievement measure of students in the same grade the year after that after having received some instructional treatment. If a t-test of the differences between means of independent samples yields a statistically significant result, and that difference is large enough to be considered educationally significant, the treatment is said to be successful.

A disadvantage to this design is that its use depends on the yearly availability of appropriate equivalent measures. Also, Cook and Cambell (1979) indicated that the design in this form is weak because differences between the two groups (other than the treatment) could cause a difference in achievement (selection). Events other than the treatment could depress achievement for the control cohort or enhance it for the treatment cohort (history). Variations in testing conditions or procedures could also be responsible for differences in performance. The design can be strengthened by adding a pretest if it is available. This augmented design is shown as follows:



The pretest could be an end-of-year test or one that is routinely

administered in the fall. The data would be analyzed using regression analysis, the same method used in the regression-discontinuity design. Here, however, each treatment student would be given the "dummy" code "1" and each student in the cohort from the year before a code of "0."

Example of cohort design use. Garretson, Vertuno, Roblyer and King (1983) used this design in a study of a computer-based reading and writing program, the "Writing to Read" program from IBM, Inc. Approximately fifty students from each grade served as subjects, and it was not considered feasible to withhold the treatment from any of them. A readiness test was available for both the treatment and control kindergarten cohort, so it was used as a pretest. No pretest measure was available for the first grade students. The dependent variable was the reading score (prereading for kindergarten) from the California Test of Basic Skills (CTBS) administered at the end of both kindergarten and first grade. The analysis for the kindergarten students was performed as follows:

STEP 1: Calculate means and standard deviations. The means and standard deviations of both groups on both measures were determined:

	Treatment (1982-83)		Control (1981-82)	
	Mean	SD	Mean	SD
Reading	129.00	13.73	124.59	14.13
Pretest	6.29	1.47	6.11	1.57

STEP 2: Perform regression analysis. This analysis actually involves several steps, but they will not be detailed here since the SPSS program described in Appendix D accomplishes them. The results were as follows:

Variable	b	F	Significance
Treatment (b1)	3.79	2.12	>.05
Pretest (b2)	4.10	22.54	<.05
Intercept (a)	99.52	314.96	<.05

Data for the first grade were analyzed using a t-test for comparing two groups. The results were:

	Treatment (1982-83)		Control (1981-82)	
	Mean	SD	Mean	SD
Reading	63.79	14.36	61.56	18.80

t-ratio = .65
(p>.05)

The results for neither grade showed a significant treatment effect. These findings are not in agreement with those in the value-added analysis in which the treatment effect for the

Nonsense Words Tests were significant. The different results could be due to a number of reasons. (See discussion in Garretson et al.)

However, despite its failure to demonstrate a significant treatment effect, the cohort design is considered to be superior to a design in which the control group is not equivalent to the experimental group (e.g. the groups are obtained from different schools). The populations from which the cohorts were drawn were probably more similar than those using students from different schools. Also, the school environment (principal, teachers, facilities, etc.) for the cohorts would probably be more similar than would the environments in different schools.

DISCUSSION AND SUMMARY

Although most research studies reported in the educational technology literature use a non-treatment control group design, this approach is often not feasible in non-laboratory situations. The microcomputer movement in schools has created a wealth of new research opportunities, many of which will require designs more suited to actual classroom implementation. Five such designs are described here:

1. Sequential analysis
2. Value-added analysis
3. Non-equivalent dependent variables
4. Regression-discontinuty
5. Cohort design

Depending upon the needs of the researcher and the characteristics of the research situation, each of these can serve as a practical alternative to a design requiring a non-treatment control group. To aid school researchers in employing these designs, statistical analysis methods and tools are also provided for each one. Several caveats are, however, in order.

Choosing the Appropriate Design

Each of the designs described here can be appropriate for studying various aspects of computer-aided instruction. But like all research designs, each has assumptions which must be met and requirements which must be considered if the research is to be perceived as valid. The flowchart shown in Figure 5 summarizes a decision path which one might take to choose the most appropriate design for the situation.

[Figure 5 about here.]

However, even after selecting the most useful of these designs, the researcher must recognize that, like all designs, each has its strengths and weaknesses. Using two of these designs together can work to make the total study stronger. For example, in Garretson et al (1983), the cohort design and the value-added analysis were used. Ideally, the results of these two designs would be the same and thus support the conclusions of the research.

Specifying Power

The sequential analysis method is the only one of the five which forces the researcher to set alpha, Beta and effect size ahead of time. However, this practice is desirable for all designs (including true experiments) in order to determine that the study has sufficient power to reject the null hypothesis when it is false or to allow acceptance when it is true (Cohen, 1969).

Required Research Expertise

Several aids are provided here to assist those with limited statistical experience in selecting and implementing the most appropriate of these designs. It should be apparent, however, that using these designs require some expertise in both research and statistics. The minimum training would probably include introductory courses in research methods, descriptive statistics and inferential statistics. And, of course, some experience in implementing research studies is helpful. To those with such expertise available to them, the designs described here can and should be useful in implementing research to direct the course of their educational computing activities.

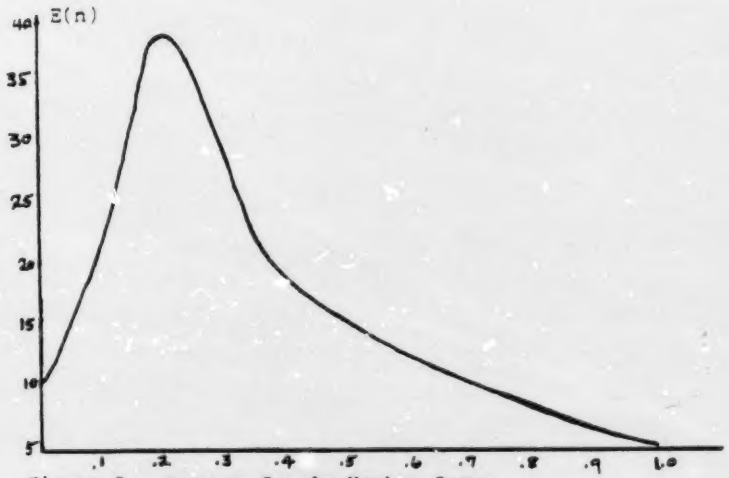


Figure 2. Average Sample Number Curve

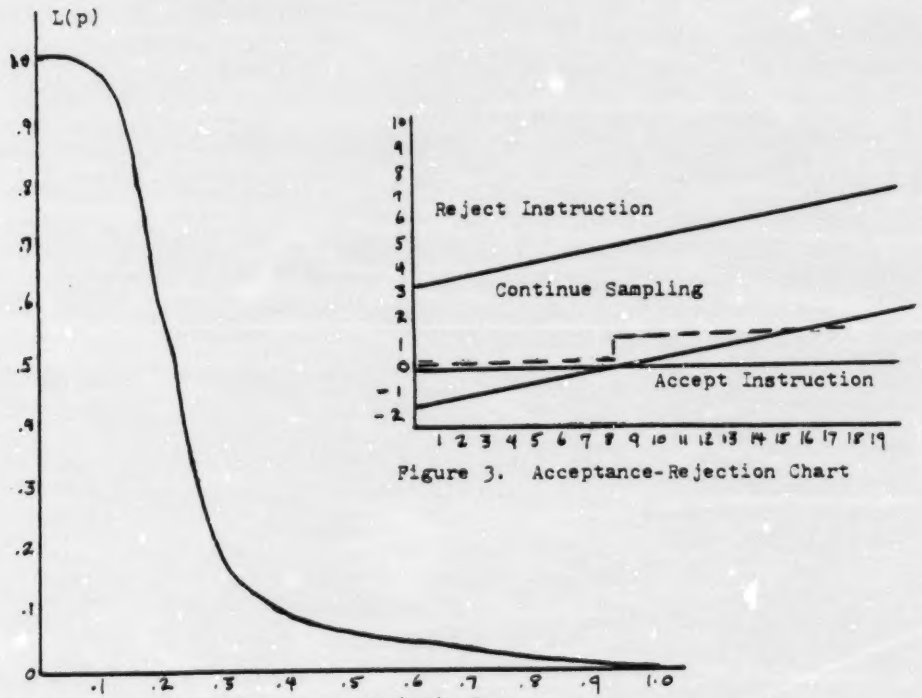


Figure 1. Operating Characteristic Curve

Figure 3. Acceptance-Rejection Chart

BEST COPY AVAILABLE

Obj.	Pretest	Interim Tests			Posttest
~	1	2	3	4	
A	O_{a1} X_a	O_{a2}	O_{a3}	O_{a4}	O_{am}
B	O_{b1}	O_{b2} X_b	O_{b3}	O_{b4}	O_{bm}
C	O_{c1}	O_{c2}	O_{c3} X_c	O_{c4}	O_{cm}
D	O_{d1}	O_{d2}	O_{d3}	O_{d4} X_d	O_{dm}
.	.				
.	.				
J	O_{j1}	O_{j2}	O_{j3}	O_{j4}	O_{jm}

Figure 4. Multiple Waves, Multiple Variables Design (Berquist and Graham, 1980)

BEST COPY AVAILABLE

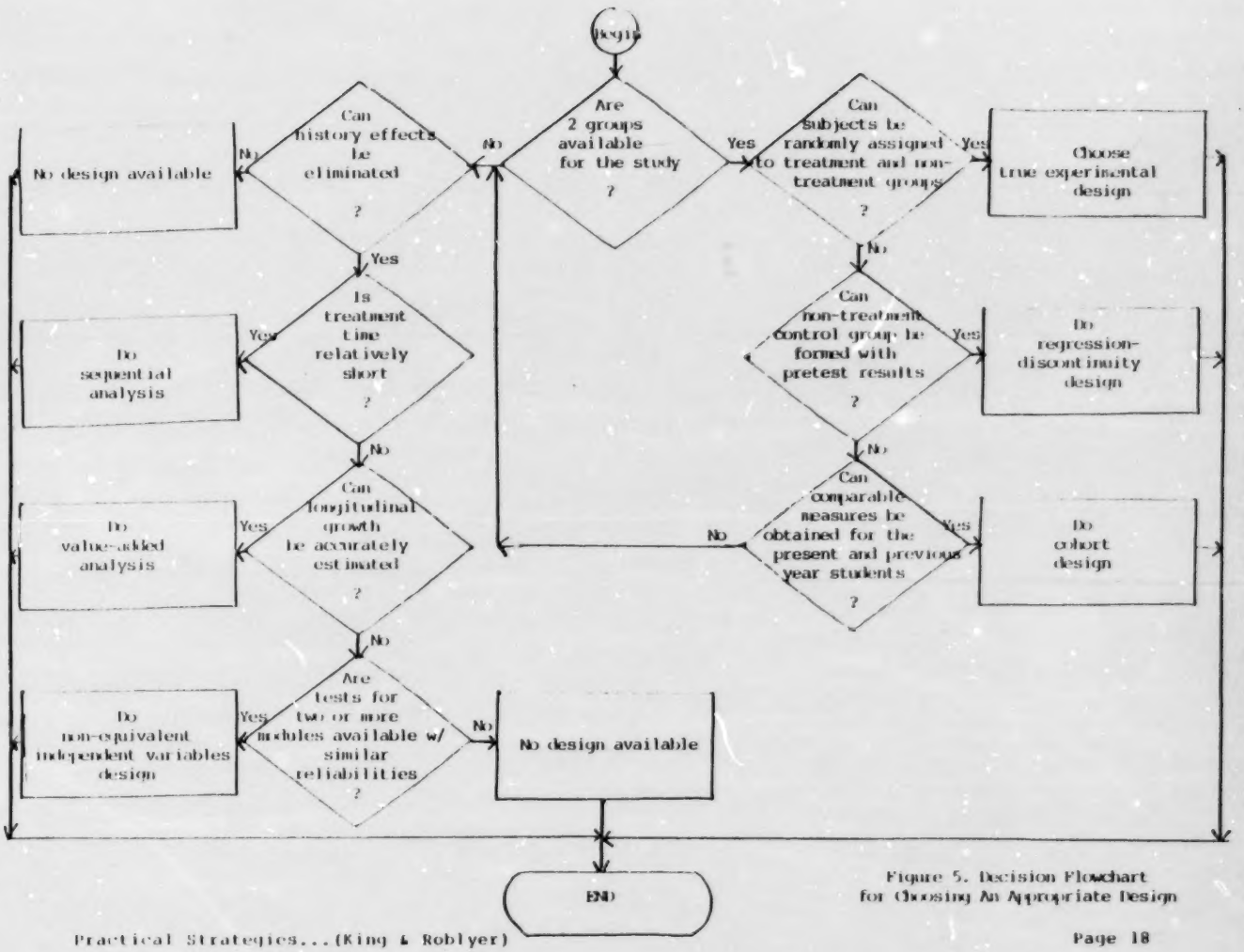


Figure 5. Decision Flowchart for Choosing An Appropriate Design

Practical Strategies...(King & Roblyer)

Page 18

BEST COPY AVAILABLE

References

Alessi, S.M., Siegel, M, Silver, D., & Baines, H. Effectiveness of a computer-based reading comprehension program for adults. Journal of Educational Technology Systems, 1982-83, 11 (1), 43-5 .

Berquist, C.C., & Graham, D.L. Developing models for special education. Using a multiple time-series, multiple baseline design in the ESEA Title IV project. Evaluation Review, 1980, 4 (3),307-321.

Bryk, A.S., & Strenio, J.F., & Weisberg, H.I. A method for estimating treatment effects when individuals are growing. Journal of Educational Statistics, 1980, 5 (1), 5-34.

Bryk, A.S., & Weisberg, H.I. Value-added analysis: A dynamic approach to the estimation of treatment effects. Journal of Educational Statistics, 1976, 1 (2),127-155.

Burr, I.W. Statistical quality control methods. New York: Marcel Dekker, Inc., 1976.

Campbell, D.T., & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally College Publishing Company, 1963.

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1969.

Cook, T.D., & Campbell, D.T. Quasi-experimentation designs and analysis issues for field settings. Chicago: Rand MacNally College Publishing Company, 1979.

Crandall, N. An analysis of the impact of CAI on a program designed to ameliorate the effects of racial isolation in the Los Nietos School District. Washington, D.C.: Office of Education, 1977. (ERIC Document Reproduction Service No. ED 147 395)

Epstein, K. I., & Knerr, C.S. Applications of sequential testing procedures for performance testing. In D.J. Weiss (Ed.) Proceedings of the 1977 Computerized Adaptive testing Conferences. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Fletcher, J.D., & Atkinson, R.C. Evaluation of the Stanford CAI program in initial reading. Journal of Educational Psychology, 1972, 63 (6),597-602.

Garretson, P., Vertuno, E., Roblyer, M., & King, F.J. A computer-based reading and writing program for K-1. Presentation at the Florida Instructional Computing Conference, Tampa, 1983.

Holland, P.W. Computer-assisted instruction: A longitudinal study. Panel presentation at the Annual Conference of the American Educational Research Association, Boston, April 1980.

Huitema, B.F. The analysis of covariance and alternatives. New York: John Wiley and Sons, 1980.

Jones, M. TICCIT applications in higher education: Evaluation results. Paper presented to the Association for Development of Computer-based Instructional Systems, Dallas, March 1978. (ERIC Document Reproduction Service No. ED 107 039)

Lysiak, F., Wallace, S., & Evans, C. Computer-assisted instruction 1976-76 evaluation report. Fort Worth: Fort Worth Independent School District, 1976. (ERIC Document Reproduction Service No. ED 140 495)

Mosteller, F, & Tukey, S.W. Data analysis and regression. Reading, Massachusetts: Addison-Wesley, 1977.

Poore, J.H., Qualls, J.E., & Brown, B.L. The educational effectiveness and economics of delivery of the PLATO basic skills mathematics lessons: A field study. AEDS Journal, 15 (1), 31-51.

Ragosta, M., Holland, P.W., & Jamison, D.P. Computer-assisted instruction and compensatory education: The ETS/LAUSD study. The final report, #19. ETS: Princeton, New Jersey, 1982.

Reichardt, C.S. The statistical analyses of data from non-equivalent group designs. In T.D. Cook and D.T. Campbell Quasi-experimentation design and analysis: issues for field settings. Chicago: Rand McNally College Publishing Company, 1979.

Suppes, P., & Morningstar, M. Computer-assisted instruction at Stanford, 1966-68: Data, models, and evaluation of the arithmetic programs. New York: Academic Press, 1972.

Wald, A. Sequential analysis. New York: John Wiley and Sons, 1947.

Appendix A
Applesoft Basic Program for Sequential Analysis

The program requires the user to enter proportions for the AQL and RQL, as well as values for alpha and Beta. H1, H2, and S are used to construct the upper (UL) and lower lines (LL) of the acceptance-rejection graph:

LL = -H1 + Sn
UL = H2 + Sn

where n is the nth subject. D1 and D2 in the output are values for the lower and upper lines where n=10. They are given to facilitate plotting of the lines.

Following the listing of the truncation number and input specifications, the user is asked whether or not a table of values should be printed. If the answer is "yes," the acceptance/rejection table is printed. It is simply a tabular representation of the graph.

Coordinates for the Operating Characteristics (OC) Curve and the Average Sample Number (ASN) Curve are printed if the user desires them. In both curves, p (the true proportion of failures) is plotted on the horizontal axis. L(p), the probablity of rejecting the instruction, and E(N), the average sample numbers, are plotted on the vertical axis.

The following sets of numbers can be used to check your program entry:

Input

	AQL	RQL	alpha	Beta
(1)	.90	.70	.01	.10
(2)	.90	.70	.05	.15
(3)	.85	.75	.05	.15

Output

	H1	H2	S	Truncate
(1)	-1.698	3.333	.186	112
(2)	-1.367	2.099	.186	57
(3)	-2.902	4.455	.197	245

```
15  INPUT "ENTER PROPORTION FOR H
    IGH QUALITY ";P1
20  PRINT
25 P1 = 1 - P1
30  INPUT "ENTER PROPORTION FOR L
    OW QUALITY ";P2
35  PRINT
40 P2 = 1 - P2
45  INPUT "ENTER ALPHA ";A1
46  PRINT
50  INPUT "ENTER BETA ";B1
51  PRINT
55  INPUT "ARE ALL ENTRIES CORREC
    T? Y^N ";N$
56  PRINT
60  IF N$ = "N" GOTO 15
70 G1 = ( LOG (P2 / P1)) * .43429
    448
75 G2 = ( LOG ((1 - P2) / (1 - P1
    ))) * .43429448
80 A = ( LOG ((1 - B1) / A1)) * .
    43429448
85 B = ( LOG (B1 / (1 - A1))) * .
    43429448
90 H1 = B / (G1 - G2)
95  PRINT "H1=INTERCEPT AT ZERO F
    OR THE LOWER LINE=  ";H1
96  PRINT
100 H2 = A / (G1 - G2)
105  PRINT "H2=INTERCEPT AT ZERO
    FOR THE UPPER LINE= ";H2
106  PRINT
110 G3 = G1 - G2
115 G4 = ( LOG ((1 - P1) / (1 - P
    2))) * .43429448
120 S = G4 / (G1 - G2)
122  PRINT "S = ";S
123  PRINT
125 D1 = H1 + (S * 10)
130  PRINT "D1=LOWER LINE VALUE A
    T N EQUALS 10 =     ";D1
131  PRINT
135 D2 = H2 + (S * 10)
140  PRINT "D2=UPPER LINE VALUE A
    T N EQUALS 10 =     ";D2
141  PRINT
150 N = (3 * (A * B)) / (G1 * G2)

155  PRINT "TRUNCATE AT N= ";N
156  PRINT
160  PRINT "WALD,""P1="1 - P1","""
    P2="1 - P2","""ALPHA="A1","""B
    ETA="B1
161  PRINT
166  PRINT
170  INPUT "DO YOU WANT TO PRINT
    A TABLE OF VALUES? Y ^ N:  "
    ;N$
171  PRINT
175  IF N$ = "N" GOTO 250
180  PRINT "N","ACCEPT","REJECT"
181  PRINT
185 N = 0
190 I = 10
195 N = N + 1
200 R0 = (B / G3) + (N * (G4 / G3
    ))
210 R1 = (A / G3) + (N * (G4 / G3
    ))
220  PRINT N TAB( 16)R0 TAB( 31)R
    1
225  IF N = I GOTO 235
230  GOTO 195
231  PRINT
235  INPUT "DO YOU WANT ANOTHER 1
    0 VALUES? Y ^ N:  ";N$
236  PRINT
240  IF N$ = "N" GOTO 250
245 I = I + 10
246  GOTO 195
250  INPUT " DO YOU WANT A TABLE
    OF VALUES OF THE COORDINATES
     TO PLOT AN OC CURVE? Y^N: "
    ;N$
260  PRINT
270  IF N$ = "N" GOTO 340
275  PRINT
280  PRINT "COORDINATES FOR OC CU
    RVE"
285  PRINT
290  PRINT "VALUES OF P","L(P)"
295  PRINT
300  PRINT 0,1
305  PRINT
310  PRINT P1,1 - A1
314  PRINT
315 K =  INT ((S + .00005) * 1000
    0) / 10000
316 K1 = H2 / (( ABS (H1) + H2))
320  PRINT S,K1
321  PRINT
322  PRINT P2,B1
325  PRINT
330  PRINT 1,0
335  PRINT
340  INPUT "DO YOU WANT A TABLE O
    F VALUES FOR THE COORDINATES
     TO PLOT ASN CURVE? Y^ N: ";
    N$
350  IF N$ = "N" GOTO 440
360  PRINT
370  PRINT "COORDINATES FOR ASN C
    URVE"
375  PRINT
380  PRINT "VALUES OF P","E(N)"
385  PRINT
386 K2 =  ABS (H1) / S
390  PRINT 0,K2
395  PRINT
397 K4 = ((1 - A1) *  ABS (H1) -
    A1 * H2) / (S - P1)
400  PRINT P1,K4
405  PRINT

406 K5 = ( ABS (H1) * H2) / (S *
    (1 - S))
410  PRINT S,K5
415  PRINT
416 K6 = (((1 - B1) * H2) - (B1 *
     ABS (H1))) / (P2 - S)
420  PRINT P2,K6
425  PRINT
426 K7 = H2 / (1 - S)
430  PRINT 1,K7
435  PRINT
530  END
```

BEST COPY AVAILABLE

Practical Strategies... (Roblyer & King)

```
Enter proportion for high quality .9Ø
Enter proportion for low quality  .7Ø
Enter ALPHA .Ø1
Enter BETA  .1Ø
Are All entries correct? Y. N Y
H1=Intercept at Zero for the lower line= -1.69826ØØ8
H2=Intercept at Zero for the Upper Line= 3.3333733
S = .186188942
D1=Lower Line Value at N Equals 1Ø =     .163433334
D2=Upper Line Value at N Equals 1Ø =    5.195Ø8772
Truncate at N= 112.Ø9Ø667
Wald, P1=.9,P2=.7,ALPHA=.Ø1,BETA=.1

Do you want to print a table of Values? Y N: N
Do you want a Table of Values of the Coordinates to plot an
do curve? Y N: Y

Coordinates for oo Curve


Values of P    L(P)
Ø              1
.1             .99
.186188942     .662482552
.3             .1
1              Ø

Do you want a table of Values for the Coordinates to Plot ASN
Curve Y N: Y
Coordinates for ASN Curve


Values of P    E(N)
Ø              9.12217724
.1             19.1246365
.186168942     37.3635557
.3             24.8632439
1              4.Ø9590328
```

Appendix B
Computations for Value-added Analysis

1. Computation of the value added by the treatment (V)

$$V = \bar{y}_2 - \bar{y}_1 - b(\bar{d})$$

where $\bar{y}_1$ and $\bar{y}_2$ are pre and post test means

b is the slope coefficient for the regression of pretest on age, and

$\bar{d}$ is the mean time span between the pre and post tests.

2. Computation of pseudo-values (V_{*i})

A. Compute the regression coefficient (b) using the whole data set.

B. Compute a regression coefficient (b_i) with observation i removed from the data. N coefficients will be computed.

C. Compute a pseudo-value (V_{*i}) for each individual:

$$V_{*i} = y_i(t_2) - y_i(t_1) - b_{*i}\,[a_i(t_2) - a_i(t_1)]$$

where $y_i(t_2)$ and $y_i(t_1)$ are post test and pre test scores for individual i

$a_i(t_2)$ and $a_i(t_1)$ are the ages of individual i at post test and pre test times, and

b_{*i} is computed as shown below:

$$b_{*i} = Nb - (N - 1)\,bi$$

D. Compute the mean and standard error of the V_{*i}'s.

Calculate a t-ratio by dividing the mean by the standard error.

The data file on the next page contains scores for age (in months), and pre test and post test scores for 10 subjects. The first SPSS procedure file computes the regression coefficient and the sums, sums of squares, and sum of cross products needed for the next program.

Appendix B (cond.)

The output of the procedure is as follows:

Regression coefficient (b)	1.134
Sum of age scores (SX)	690
Sum of pre test scores (SY)	224
Sum of cross products of age and pretest (SXY)	15778
Sum of squares of age (SXQ)	47894

The output of the second SPSS procedure gives the mean (2.028) and standard error (2.515) of the pseudo-values (V_{*i}).

The t-ratio is 2.028/2.515 = .806. Thus, for these 10 cases, the value added by the treatment is not statistically significant.

DATA FILE

```
73 25 29
72 38 39
73 30 35
79 30 44
63  7 12
69 10 16
69 22 27
68 22 36
64 28 30
60 12 19
```

SPSS PROCEDURE #1

```
 5.     VARIABLE LIST
 5.005  AGE PRE POST
10.     INPUT FORMAT
10.005 FIXED(F2.0,2F3.0)
15.     COMPUTE
15.005 SXY=AGE*PRE
16.     COMPUTE
16.005 SXQ=AGE**2
20.     REGRESSION
20.005  VARIABLES=AGE PRE/REGRESSION=PRE WITH AGE
30.     CONDESCRIPTIVE
30.005 AGE,PRE,POST,SXY,SXQ
```

SPSS PROCEDURE #2

```
 5.     VARIABLE LIST
 5.005  AGE PRE POST
10.     INPUT FORMAT
10.005 FIXED(F2.0,2F3.0)
50.     COMPUTE
50.005 SXY=15778
55.     COMPUTE
55.005 SXQ=47894
70.     COMPUTE
70.005 SX=590
75.     COMPUTE
75.005 SY=224
80.     COMPUTE
80.005 B=1.134
85.     COMPUTE
85.005 NUM=(9*(SXY-(AGE*PRE)))-((SX-AGE)*(SY-PRE))
90.     COMPUTE
90.005 DENOM=(9*(SXQ-(AGE**2)))-((SX-AGE)**2)
92.     COMPUTE
92.005 BI=NUM/DENOM
94.     COMPUTE
94.005 BSI=(10*B)-(9*BI)
S/C/RETURN
 C
 96.     COMPUTE
 96.005 V=POST-PRE-(BSI*4)
 98.     CONDESCRIPTIVE
 98.005 V
 99.     STATISTICS
 99.005 1,2
```

BEST COPY AVAILABLE

Appendix c

Analysis and Results of the Regression-Discontinuity Design

The data file is shown in the upper left-hand corner of the next page. There are eight treatment students below the pretest cutting score of 40, and eight control students above it. Columns 1 and 2 contain a student ID number, Column 3 is the group membership variable (treatment=1, control=2), the next two columns contain the pre test scores, and the post test is in the last two columns.

The SPSS procedure file is given to the right of the data file. Lines 5.0 and 10.0 specify the names of the variables and the input format, respectively. Line 200 is the regression procedure. It specifies the dependent variable is the post test, and the independent variable is the treatment variable and the pretest. Line 25.0 causes means, standard deviations, and correlations to be printed.

The output of the procedure consists of means, standard deviations, and correlations of the three variables as labeled. The last three lines of output show the intecept and regression coefficients:

$$\overline{Y} = 35.00 + 9.72 + .37 \text{ Pre}$$

The columns labeled F and Sig. show that all effects are statistically significant. The treatment regression coefficient (b1) is the adjusted difference between treatments. The results are graphically illustrated in the figure that follows the output. The difference in intecepts of the cutting score is the estimate of the treatment effect.

DATA FILE

```
 1 1 11 39
 2 1 13 44
 3 1 19 42
 4 1 25 46
 5 1 24 50
 6 1 30 42
 7 1 29 54
 8 1 18 48
 9 0 43 56
10 0 42 60
11 0 50 58
12 0 51 61
13 0 55 62
14 0 54 66
15 0 62 63
16 0 63 67
```

SPSS PROCEDURE

```
 5.     VARIABLE LIST
 5.005  ID,T,PRE,POST
10.     INPUT FORMAT
10.005  FIXED(2F2.0,2F3.0)
20.     REGRESSION
20.005  VARIABLES=T,PRE,POST/
20.006  REGRESSION=POST WITH T,PRE
25.     STATISTICS
25.005  1,2
```

OUTPUT

VARIABLE	MEAN	STANDARD DEV	CASES
T	.5000	.5164	16
PRE	36.8125	17.7058	16
POST	53.6250	4.3340	16

CORRELATION COEFFICIENTS.

A VALUE OF 99.00000 IS PRINTED
IF A COEFFICIENT CANNOT BE COMPUTED.

```
PRE          -.91507
POST         -.23830    .46642

              T         PRE
             ---------
```

DEP. VAR... POST

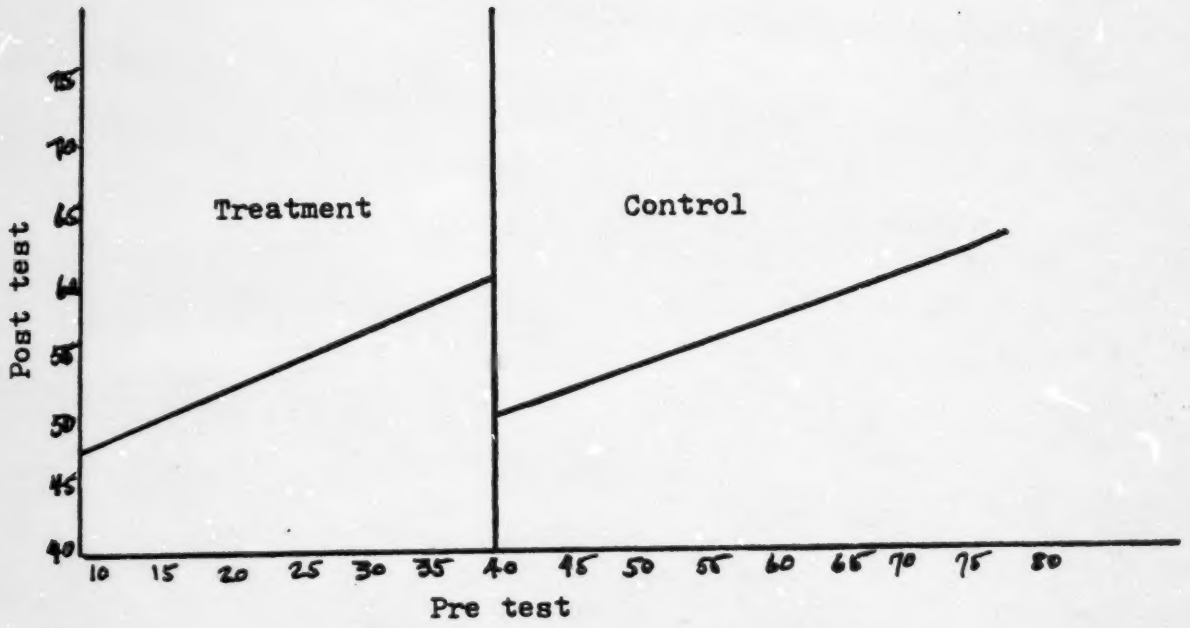
MEAN RESPONSE 53.62500 STD. DEV. 4.33397

FINAL STEP.

MULTIPLE R	.6603	ANOVA	DF	SUM SQUARES	MEAN SQ.	F
R SQUARE	.4360	REGRESSION	2.	122.847	61.423	5.025
STD DEV	3.4962	RESIDUAL	13.	158.903	12.223	SIG. .024
ADJ R SQUARE	.3492	COEFF OF VARIABILITY		6.5PCT		

VARIABLE	B	S.E. B	F	SIG.	BETA	ELASTICITY
T	9.727	4.334	5.036	.043	1.15893	.09069
PRE	.374	.126	8.741	.011	1.52692	.25857
CONSTANT	35.003	6.751	26.883	.000		

BEST COPY AVAILABLE


Treatment
Control
Post test
Pre test
40
45
50
55
60
65
70
75
80
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80